



Generative AI — The End of Systematic Reviews in PhD Projects?

By Sanaz Zamani, *Auckland University of Technology* and Roopak Sinha, *Deakin University*

In this opinion piece, we question the efficacy of students conducting systematic reviews (SRs) at the very start of their PhDs, especially now that we are riding, or drowning in, the Generative AI wave. How would the ubiquitous availability of Generative AI tools shape PhD research and supervision, and how can we capitalize on such disruptions? We report our reflections about early-stage PhD research from 10+ PhD supervisions completed by Sinha, and a recently started PhD project with Zamani as the student. We conducted a systematic literature review following traditional methods for the first four months and then revisited the task using ChatGPT and other related tools. We opine that while SRs still have value for new students, Generative AI must be incorporated, with care, to significantly enhance student learning, reinvest time savings into other, often-ignored skills, and improve the quality (breadth and depth) of SRs.

SYSTEMATIC REVIEWS IN PHD RESEARCH - CURRENT PRACTICE

Generative AI is redefining workflows in all professions, including research. Tools like ChatGPT [8] offer the ability to review large knowledge bases and deliver meaningful insights in a matter of minutes. So, in a Generative AI-equipped world, do we really need systematic reviews (SRs) in research, especially early on in PhD research?

Notably, in our own research group and in several others that we collaborate with, systematic reviews have become a big hit in PhD research. Published systematic mapping studies and literature reviews attract higher citations over the long term [1,7]. Often, in the first large project assigned to students, SRs are helpful in identifying gaps in current state-of-the-art, identifying what to work on, and developing a comprehensive understanding of a new area for researchers. In many cases,

In many cases, SRs are the first tangible bricks with which the foundation of PhD research is built. ... Regardless of who conducts them, SRs can age badly and become outdated, especially in fast-moving research areas or when the SRs cover a very narrow slice of a research area.

SRs are the first tangible bricks with which the foundation of PhD research is built.

On the downside, students can spend months working on SRs, leaving other equally important research skills dormant during the early stages of their research journeys. The traditional way of conducting SRs has a high opportunity cost, and publishing SRs in top venues requires significant additional time; the quest to publish SRs further delays the “real work” needed to complete a PhD. Consequently, it is now common to see “top-heavy” PhD theses that cover problem formulation extremely well (thanks to SRs) but are relatively light on technical contributions. On average, the quality of SRs conducted by PhD students is (expectedly) low, as they do not possess sufficient research experience to extract deep insights during their first such experience. Leading researchers and experts are often too busy to conduct and publish SRs. Regardless of who conducts them, SRs can age badly and become outdated, especially in fast-moving research areas or when the SRs cover a very narrow slice of a research area.

Generative AI tools provide speed and scale; they can access, review, and connect multiple sources with faster speed, coverage, and transparency. Generative AI-created content is already finding its way into research articles, but current uses are, at best, inconsistent and, at worst, unethical.

Over time, SRs have become increasingly difficult to publish due to increased supply. Published SRs are potentially more helpful for supervisors as they attract higher citation counts but are not so good for students; having a highly-cited secondary study on one's CV is often not beneficial for starting a post-PhD R&D career (industry or academic). One of the main reasons SRs have become popular is that they make secondary studies repeatable and agnostic to bias. While this is a noble objective, the process of conducting SRs has become overly mechanical. The de-skilling of the review process and the increased researcher fungibility impact creativity. Interestingly, this mechanization leans itself well to the automation offered by Generative AI.

We felt all of the above when Sinha completed a draft systematic literature review on using technology to improve emotional well-being during the first four months of her PhD following the traditional approach. At the time of writing, we were facing another 3-4 months of work to make the review publishable. We decided to investigate the use of Generative AI to see if it offered what, if any, improvements. We revisited the SLR process and reran all the stages using assistance from Generative AI tools. We report that experience in the next section.

GENERATIVE AI-ASSISTED SYSTEMATIC REVIEW PROCESS

Several studies, such as [1], explore multidisciplinary perspectives on the opportunities, challenges, and implications of generative conversational AI, with a specific focus on ChatGPT in research, practice, and policy. Some research, such as in [2], explores the application of AI in conducting SRs to streamline and enhance the review process and provide researchers with a quicker and more cost-effective approach while maintaining methodological rigor. Castillo-Segura et al. [3] explore the potential of generative AI in accelerating SLRs in educational technology. George [5] emphasizes the need for reform to align graduate programs with evolving workplace demands and societal needs.

Generative AI tools provide speed and scale; they can access, review, and connect multiple sources with faster speed, coverage, and transparency. Generative AI-created content is already finding its way into research articles, but current uses are, at best, inconsistent and, at worst, unethical. Consistent usage patterns for Generative AI for automating the mechanical steps in SRs can improve quality and accuracy and re-

purpose researcher time on developing and practicing other research skills.

Creating our generative-AI-assisted review involved learning from the above sources and coming up with solutions where the literature was unclear. The overall process we followed is presented here.

PHASE 1: SEARCHING

Prompt engineering through ChatGPT (v3) was used for the following.

1. *Refining research questions:* ChatGPT can help ensure questions are demonstrably better formed and more amenable for systematic reviews.
2. *Creating comprehensive inclusion and exclusion criteria:* ChatGPT could list many potential criteria that could then be refined to scope the review.
3. *Developing robust search queries:* The tool can provide several well-formed search queries that were used to run quick searches on selected scientific databases. These quick iterations helped develop a final query.

PHASE 2: SCREENING

Machine learning tools like ASReview [9] have been quite successful in screening titles and abstracts for systematic reviews and identifying relevance based on specific user-provided criteria. ChatGPT was used to improve the selection of criteria to get better results from ASReview. The pruned list of articles could then be used in the next steps for further review.

PHASE 3: DATA EXTRACTION

We used the Generative AI tool ChatPDF [6] to read screened studies and automatically respond to pre-defined questions for data extraction. The questions extracted from the research questions were processed through ChatGPT to make them more precise and to ensure that data were comprehensively collected. ChatPDF was significantly faster in identifying whether a study was relevant for the review, compared to the traditional method in which the researcher is forced to manually read each article and decide. However, it did miss some information in several cases and brought up some irrelevant information when answering some questions. For example, ChatPDF was able to respond quickly and comprehensively to find out about specific technologies used in previous research works. Sometimes, though, it ignored negations in the text and highlighted minor points rather than key concepts, leading to low-quality synthesis.

Generative AI is causing fundamental disruption in how we conduct systematic reviews. Those who adapt faster will have an advantage in the short term, but in the long term, the research community will need to develop standardized patterns for conducting reviews using Generative AI.

These issues have to be addressed through our (human) intervention, and potential time savings from using these tools could be lost in some cases. However, students potentially gain additional skills like prompt engineering, discernment, and ethical practice in engaging with Generative AI while conducting SRs, which is a beneficial learning consequence for them.

DISCUSSION

Generative AI assistance delivered significant benefits in speed and quality (comprehensiveness) when conducting systematic reviews. In our case, the overall Generative AI-based process was at least four times faster than the traditional process to produce the first complete drafts of an SR document. Students also develop skills to use these tools, and these skills are seen as important in today's world. Tools could access and analyze many research studies in a fraction of the time it takes manually. Most benefits were delivered where the review process steps are highly mechanical and/or for which specialized tools (like ASReview, ChatPDF, etc.) exist.

Some weaknesses and threats are evident also. The tools required significant human guidance to ensure quality and correctness throughout the review process. Supervisors may now be expected to master and teach prompt engineering in research contexts to their PhD students. Furthermore, while these tools can automatically pick up older (2019 and earlier) open-access research articles online, they currently cannot find articles behind paywalls. In addition to inherent bias picked up from the internet during their training, the tools show an additional bias towards open-access articles during the first few phases of conducting a systematic review. Ethical concerns with the use of Generative AI are rising, and their use in SRs is no exception. Researchers can easily skew SRs by asking targeted/biased questions and/or guiding the review to produce the results they expect (confirmation bias). Finally, the increased speed may mean students end up conducting multiple SRs during the initial phases of their PhDs, resulting in almost no savings in time overall.

From the perspective of Generative AI tool users, we are hopeful that many of the technical weaknesses in current tools will likely be phased out with improvements in technology. However, we feel that the need for significant, quality human intervention cannot be easily addressed. Human judgment is critical in formulating research questions. Generative AI can help conduct a review faster, but the question, “*Is this the right*

review to conduct?” is perhaps best answered through a Human-Generative AI partnership. Humans are also currently critical for discerning the correctness of the information being generated and are in charge of using these tools ethically. One way forward could be the development of robust, standardized processes for Generative AI-based SRs. This is beyond the scope of this short opinion piece but is a clear direction for future research for our community.

In summary, as with every other field, Generative AI is causing fundamental disruption in how we conduct systematic reviews. Those who adapt faster will have an advantage in the short term, but in the long term, the research community will need to develop standardized patterns for conducting reviews using Generative AI. ❖

References

- Alimoradi, F., Javadi, M., Mohammadpoorasl, A., Moulodid, F., Hajizadeh, M. The effect of key characteristics of the title and morphological features of published articles on their citation rates, *Annals of Library & Information Studies*, 63 (2016), 74-77.
- Atkinson, C.F. Cheap, Quick, and Rigorous: Artificial Intelligence and the Systematic Literature Review. *Social Science Computer Review*, 42, 2 (2023), 1-18.
- Castillo-Segura, P., Alario-Hoyos, C., Kloos, C.D., Panadero, C.F. Leveraging the Potential of Generative AI to Accelerate Systematic Literature Reviews: An Example in the Area of Educational Technology, *2023 World Engineering Education Forum-Global Engineering Deans Council (WEEF-GEDC)*, (2023), 1-8.
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., et al. “So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges, and implications of generative conversational AI for research, practice, and policy. *International Journal of Information Management*, 71 (2023), 277-304.
- George, A.S. The Potential of Generative AI to Reform Graduate Education, *Partners Universal International Research Journal*, 4 (2023), 36-50.
- Panda, S. Enhancing PDF interaction for a more engaging user experience in library: Introducing ChatPDF. *IP Indian Journal of Library Science and Information Technology*, 8 (2023), 20-25.
- Qu, Q., Zhang, X. Bibliometric study on the highly cited papers and correlation between citation frequency and download frequency in Chinese medicine journals, *International Journal of Pattern Recognition and Artificial Intelligence*, 10 (2021).
- Ray, P.P. ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations, and future scope. *Internet of Things and Cyber-Physical Systems*, 3 (2023), 121-154.
- Van De Schoot, R., de Bruin, J., Schram, R., Zahedi, P., de Boer, J., Weijdemans, F., et al. An open source machine learning framework for efficient and transparent systematic reviews. *Nature Machine Intelligence*, 3 (2021), 125-133.

Sanaz Zamani

Department of Computer Science and Software Engineering
Auckland University of Technology
Auckland, New Zealand
sanaz.zamani@autuni.ac.nz

Roopak Sinha

School of Information Technology
Deakin University
Melbourne, Australia
roopak.sinha@deakin.edu.au

DOI: 10.1145/3657304

Copyright held by owner/author(s).